Discounted Rewards

$State, s \in S$

$Action, a \in A(s)$

$Reward, r \in R(s, a, s')$

$Transition\ Probability, P(s, a, s')$

$Optimal\ Policy, \pi^*(s)$

$Return, G_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k}$

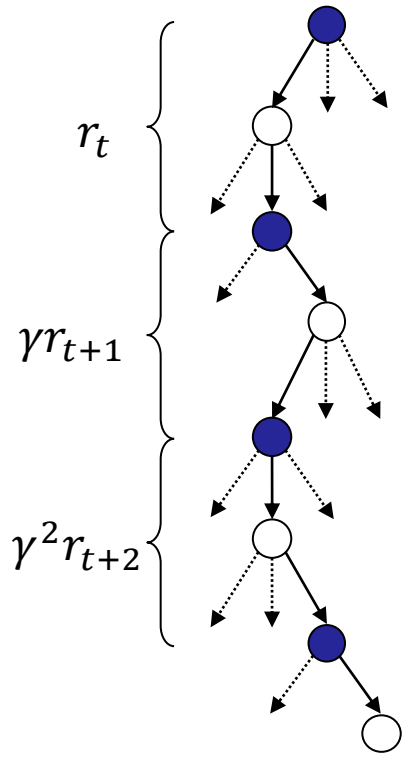
- $G_t = \text{sum}([r_t,\ r_{t+1},\ r_{t+2}, \ldots, r_T])$
- Problem: Diverging return.
- Solution:
  - Fixed step termination: **Finite horizon**
  - Guaranteed terminal state: **Absorbing State**
  - **Discounted rewards**
- $G_t = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \cdots = \sum_{k=0}^{\infty} \gamma^k r_{t+k}$

Properties of Discount Factor, γ

- $\sum_{k=0}^{\infty} \gamma^k r_{t+k} \leq \frac{r_{max}}{1-\gamma}$, $\Rightarrow$ Bounded return.
  - $\gamma \in (0, 1)$
  - Small $\gamma \Rightarrow$ More focus on present return
  - Large $\gamma \Rightarrow$ More focus on future return

This return is not sufficient, to learn optimal policies in unknown environments.

- MDP Search Trees



Future rewards have low influence on decision

https://www.profkumar.co.in/